ABSTRACT OF THE DISCLOSURE

A modeling framework for predicting the number, type, and distribution of crossovers in directed evolution experiments is disclosed. The framework provides for determining how fragmentation length, annealing temperature, sequence identity, and number of shuffled parent sequences affect the number, type, and distribution of crossovers along the length of reassembled sequences. This framework allows for the optimization of directed evolution protocols in response to a particular enzyme or protein design challenge. One method according to the present invention includes applying equilibrium thermodynamics to a plurality of sequences to determine statistics of hybridization; and parameterizing an assembly algorithm using the statistics of hybridization.